Hewlett Packard Enterprise | NVIDIA

HPE ProLiant Gen11 AI 最佳化解決方案

支援從邊緣到資料中心再到雲端的 AI 工作負載

搭上 AI 革命順風車

人工智慧 (AI) 以 20 世紀 50 年代中期完成的基礎工作為依據，已從「開創性的科幻小說」發展成為各行各業越來越多的公司使用的主流技術。由於機器學習 (ML) 和深度學習 (DL) 等 AI 工作負載需要用到大量運算資源，因此只能由高效能、高密度伺服器利用加速器和進階圖形處理單元 (GPU) 來處理。當前，對具備 AI 能力的伺服器需求正在呈指數級增長。

為了回應您對強大伺服器的需求，以滿足您的 AI 要求，HPE 和 NVIDIA® 借助業界領先的伺服器創新和全球領先的 AI 專業知識，提供了搭載新一代 NVIDIA GPU 的 HPE ProLiant Gen11 伺服器。

這些效能密集型的產業標準伺服器，具備加速業務創新和利用 AI 革命所需的可擴充性、效率和效能。

HPE ProLiant Gen11 伺服器 —

「到 2025 年，50% 的企業將設計出人工智慧 (AI) 協調平台來實作 AI，而 2020 年這個比例還不到 10%。」[1]

[1]《Our Top Data and Analytics Predicts for 2021》(我們對 2021 年的主要資料和分析預測)，Gartner 部落格，2021 年 1 月 12 日。

助力當前和未來 AI

選擇 HPE ProLiant Gen11 for AI 的原因

創新的 HPE ProLiant Gen11 伺服器提供先進的工程解決方案，可解決當今的混合雲基礎架構挑戰。它們將內部部署和雲端運算的優點與以下功能相結合：

- **直覺式**雲端作業體驗
- HPE **值得信賴的**安全設計
- **針對大型**、複雜的 AI 工作負載最佳化效能

最新動態

全新 HPE ProLiant Gen11 伺服器產品組合採用了全新的 GPU 支援方法；在某些情況下，將 GPU 移至機箱前端以改善氣流、增加 GPU 密度，並為每個 GPU 提供專用電源以延長 GPU 開機運作時間。如此一來，HPE ProLiant Gen11 伺服器便可以為高瓦數 GPU 提供動力，有效滿足各種 AI 特定工作負載的要求。

前端籠式設計，讓特定 NVIDIA GPU 能直接相互通訊。這種通訊能力，讓 GPU 的可用記憶體得以合併，從而提高資料交換的速度。結果就是效能顯著提高，能處理具有 1000 多億個參數的 AI 模型。

此外，HPE 伺服器還搭載第四代 Intel® Xeon® 可擴充處理器或第四代 AMD EPYC™ 9004 系列處理器。這些伺服器都提供了廣泛的功能，可最佳化功耗和效能並最大化 CPU 資源，協助您的組織實現永續性目標。

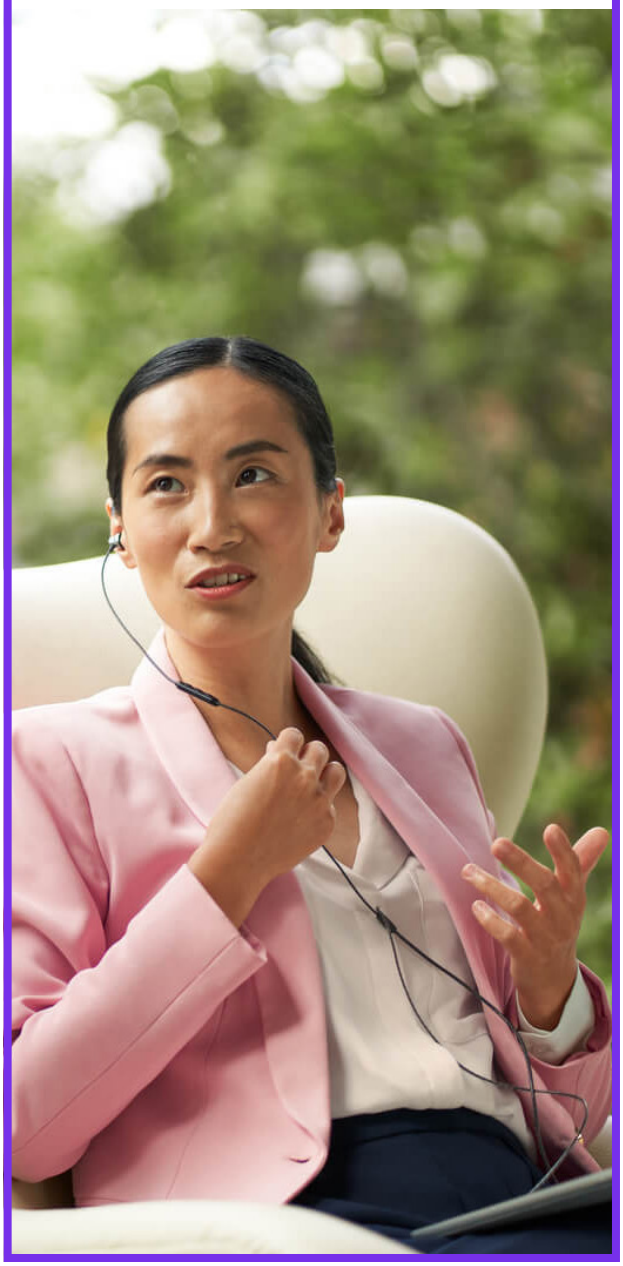
所有 HPE ProLiant 伺服器都內建 HPE iLO 6。透過 HPE iLO 6，您能夠從世界任何地方安全地設定、監控及遠端管理 HPE ProLiant 伺服器，並提供有關 NVIDIA GPU 的寶貴資產資訊，例如裝置庫存、溫度報告、電源管理、韌體報告、健全狀況及事件記錄。

新一代 HPE ProLiant Gen11 伺服器的管理已有新的突破。利用 HPE GreenLake 邊緣到雲端平台 (包括其內建的管理應用程式和架構)，直覺式雲端操作體驗可簡化並保護從邊緣到雲端的一切作業。透過自動執行載入、更新、管理及監控等關鍵生命週期工作，HPE 伺服器透過瀏覽器型的統一介面，為任何位置的運算裝置帶來敏捷性和效率。

資料優先 GPU 現代化

搭載 NVIDIA 新一代 GPU 的 HPE ProLiant Gen11 伺服器，為當今的 AI、ML 和 DL 工作負載提供充沛動力。

- **NVIDIA L40S Tensor Core GPU**：L40S GPU 具有強大 AI 運算所需的多工作負載效能以及一流的圖形和媒體加速功能，其設計能為新一代資料中心工作負載提供充沛動力。
- **NVIDIA L4 Tensor Core GPU**：極具成本效益、薄型、低功耗、單寬尺寸，可在任何伺服器中實現高輸送量和低延遲。
- **NVIDIA L40 GPU**：雙寬加速器專為視覺運算和圖形密集型資料處理而設計。





瞭解 HPE ProLiant Gen11 AI 伺服器

表 1. HPE ProLiant Gen11 AI 伺服器概覽

			
	HPE ProLiant DL320 Gen11 伺服器	**HPE ProLiant DL380a Gen11 伺服器**	**HPE ProLiant DL385 Gen11 伺服器**
外觀尺寸	1U，單插槽	2U，雙插槽	2U，雙插槽
處理器	第 4 代 Intel Xeon 可擴充處理器	第 4 代 Intel Xeon 可擴充處理器	第 4 代 AMD EPYC 9004 系列處理器
記憶體	2 TB DDR5	3 TB DDR5	6 TB DDR5
儲存	10 個 SFF SSD	8 個 EDSFF E3.S 1T NVMe SSD	36 個 EDSFF E3.S 1T NVMe SSD
GPU	最多 2 個雙寬或 4 個單寬 GPU	最多 4 個雙寬或 8 個單寬 GPU	最多 4 個雙寬或 8 個單寬 GPU
連線	ConnectX-7 系列網路	ConnectX-7 系列網路	ConnectX-7 系列網路
管理	HPE GreenLake for Compute Operations Management (訂閱)		

完美搭配，一目了然

在當今資料優先的世界裡，可供選擇的伺服器和 GPU 選項琳瑯滿目，資料中心經理不斷尋找下一個實現 IT 基礎架構現代化的最佳方法。管理人員必須考慮精密的 AI 演算法的複雜性，以及在邊緣和雲端執行的其他資料密集型運算需求。當嘗試將工作負載需求與伺服器/GPU 解決方案做出最佳搭配時，決策過程變得更加錯綜複雜。HPE 和 NVIDIA 可以提供建議的產品配對組合，無論 AI 之旅走向何方，都能協助您輕鬆駕馭。

經 NVIDIA 認證的 HPE 系統可將您的決策提升到全新境界，為您提供參考設計來打造可擴充的加速運算單元。NVIDIA 認證可確保各種 AI 工作負載的最佳效能、可靠性和擴充規模。

NVIDIA 認證的系統

您可以放心地購買 HPE AI 解決方案，因為每個系統都經過 HPE 認證並提供：

- 打造加速系統的參考設計指南
- NVIDIA 網路可確保從儲存到 GPU 運算的超快速資料傳輸
- 使用 HPE GreenLake for Compute Ops Management，無論各個系統位於何處 (邊緣、主機託管位置還是資料中心)，都能對其進行管理

為技術找到用武之地

以下配對的技術解決方案範例可以協助您為特定的 AI 相關業務目的/工作負載選擇最佳伺服器和 GPU 組合。以下重點內容可讓您更深入瞭解每組配對的特性和功能。

表 2. HPE 和 NVIDIA 配對技術解決方案

組合	目的/工作負載	功能亮點
邊緣的電腦視覺 AI HPE ProLiant DL320 Gen11 伺服器 NVIDIA L4 GPU	• 電腦視覺和智慧影像分析 • 損失預防 • 智慧空間	• 即時電腦視覺推論非常適合邊緣 • 小巧 1U 外形的 HPE ProLiant DL320 • 多達 4 個 NVIDIA L4 GPU • 搭載 NVIDIA Metropolis • 非常適合零售、餐旅和製造業
生成式視覺 AI HPE ProLiant DL380a Gen11 伺服器 NVIDIA L40 GPU	• 3D 動畫和算圖 • 影片內容創作 • 影像產生	• HPE ProLiant DL380a 支援 4 個 NVIDIA L40 GPU • NVIDIA AI Enterprise 支援市場領先的軟體 • 增強 AI 影片效能，顯著提升個人化內容 • 非常適合媒體和娛樂、醫療保健和製造業
自然語言處理 AI HPE ProLiant DL380a Gen11 伺服器 NVIDIA L40S GPU	• NLP • 語音 AI • 詐騙偵測 • 預測性維護	• 經過 AI 最佳化的 HPE ProLiant DL380a，配備 4 個 L40S GPU • 專為驅動進階 NLP 而設計，以用於詐騙偵測和語音 AI 等使用案例 • 開發及部署微調模型 • 非常適合 FSI、製造、客戶服務業

附註：以上絕非完整的清單。請聯絡您的 HPE 或 NVIDIA 代表，討論您的具體 AI 要求。

NVIDIA 提供加速軟體

NVIDIA AI Enterprise 軟體可加速資料科學流程並簡化生產級 AI 的開發，包括生成式 AI、電腦視覺、語音 AI 等。

NVIDIA AI Enterprise 擁有 100 多個架構、預先訓練模型和開發工具，可以協助您的企業加速利用 AI 的優勢，同時簡化 AI，讓每個企業都能使用這項技術。與經過 NVIDIA 認證的 HPE 系統結合使用時，NVIDIA AI Enterprise 可確保您獲得適當水準的效能、可擴充性和企業級支援。

為 AI 成功做好準備

決策考量重點

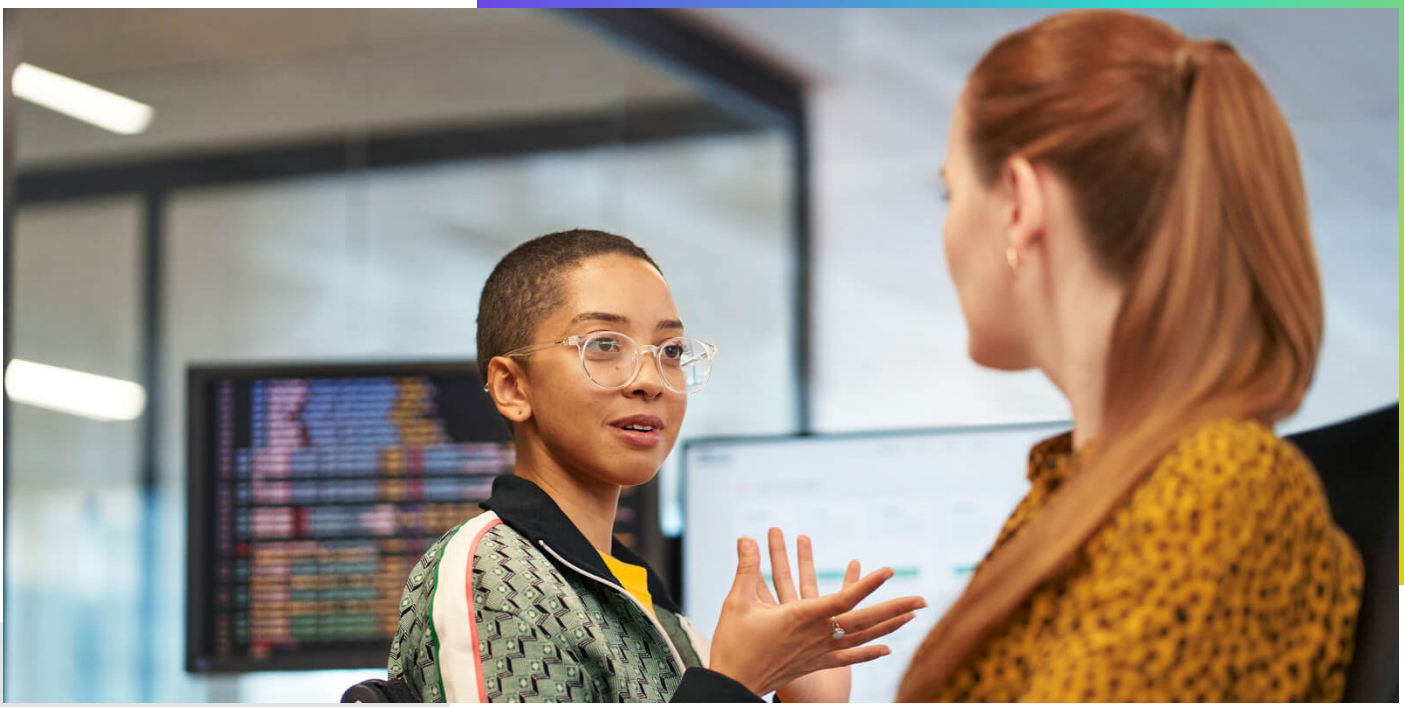
HPE 和 NVIDIA 的聯合解決方案開闢了一條通往 AI 成功的道路，使您能夠從 IT 現代化專案中獲得最大價值。無論是個別 HPE 伺服器、NVIDIA GPU 還是配對組合，每個 AI 基礎架構決策都應考量以下因素：

- **可擴充性：**選擇易於升級的伺服器，以便您可以根據 AI 處理需求的改變而增減 GPU。部署可擴充解決方案，以滿足您從邊緣到資料中心再到雲端的需求。
- **效能：**選擇的解決方案必須能執行大型複雜模型並有效處理日益具有挑戰性的 AI 工作。
- **連線：**實作提供高頻寬連線的解決方案，以便在 GPU 之間實現平穩的資料傳輸，進而支援平行處理並減少使用多個 GPU 訓練 AI 模型所需的時間。
- **管理：**每台 HPE ProLiant Gen11 伺服器都內含 HPE GreenLake for Compute Ops Management，這項工具旨在簡化及自動化整個伺服器生命週期的操作，無論您的運算基礎架構位於何處。

利用 HPE GreenLake 雲端服務

HPE GreenLake 平台可提供值得信賴的企業級雲端體驗，以加速內部部署、資料中心或主機託管位置的資料現代化計畫。透過持續監控和按需調整容量大小的靈活架構隨需擴充，在獲得洞察並最佳化營運的同時，牢牢掌控您的資料。

如需進一步瞭解 HPE GreenLake，請造訪 hpe.com/tw/en/greenlake.html。



重新想像未來

邊緣視訊分析、大型語言模型 (LLM)、NLP、AI 導向型推薦和即時搜尋 (包括 ChatGPT 和 Microsoft Bing) 等應用，已經在我們的數位生態系統中實現。由於突破性的 AI 創新，在不久的將來，自動駕駛汽車、機器人、智慧城市和智慧家居等，將會司空見慣 — 現在我們才剛開始。

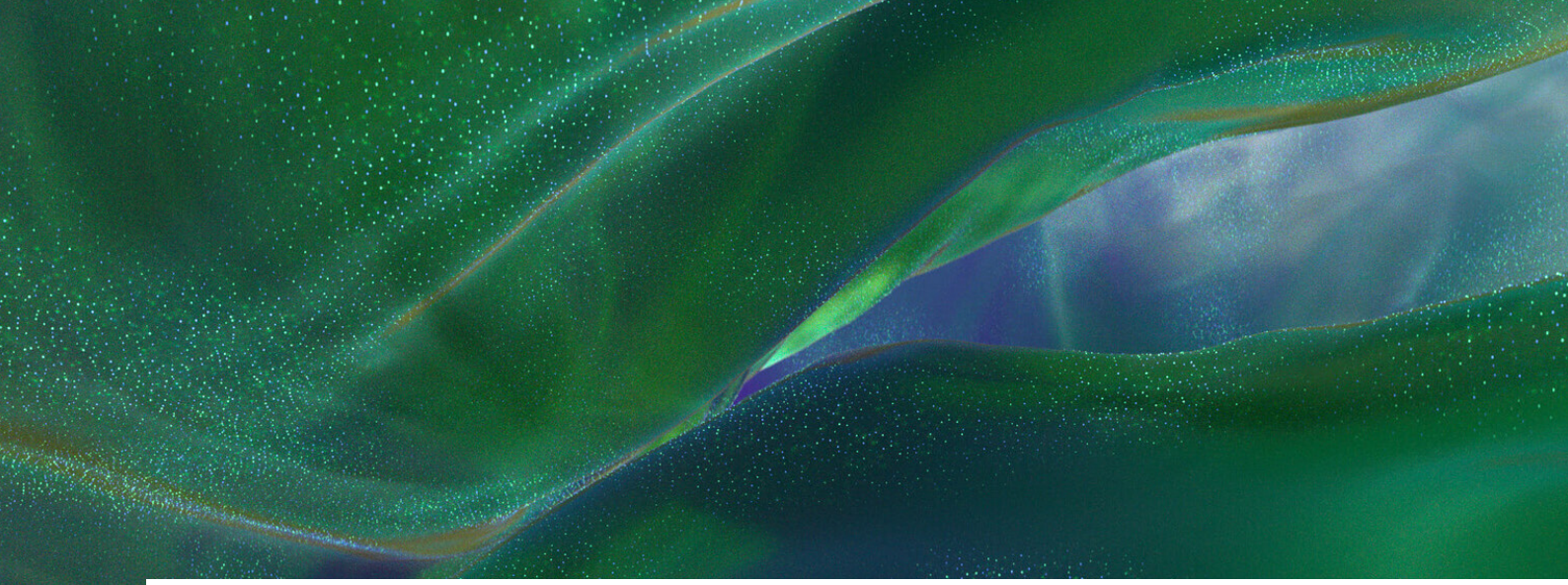
AI 的進步勢必推動變革

使用案例範例

如今，越來越多的產業使用 AI 及其他新興技術 (例如深度學習和機器學習) 來回答問題，並揭露了幾年前被認為無法解決或不可能的洞察。AI 日益成為推動進步、改善我們日常生活 (工作、家庭和娛樂) 的技術引擎。

欲探索 HPE 和 NVIDIA 在哪些領域為 AI 世界帶來重大影響，請造訪以下網頁：

- 臨床研究
- 金融服務
- 醫療保健與生命科學
- 製造業
- 體育館



HPE 與 NVIDIA 強強聯手，超越競爭對手

HPE ProLiant Gen11 伺服器和 NVIDIA 新一代 GPU 站在 AI 技術的最前端，在恰當的時間提供正確的解決方案，推動進步之輪持續向前。透過持續的研究和協作，我們的運算產品和解決方案將繼續與 AI 和未來的發展並駕齊驅。

現在，正是利用業界領先、經過 NVIDIA 認證的 HPE ProLiant Gen11 伺服器來確保您的 IT 基礎架構面向未來的最佳時機。這些創新系統可以透過加快 AI 模型和高效能資料分析的開發和部署，協助您的組織利用 AI 革命所帶來的好處。無論您的 AI 工作負載多麼複雜或要求多麼嚴苛，您都可以從 HPE 和 NVIDIA 找到合適的伺服器/GPU 解決方案。

詳情請造訪 hpe.com 的 NVIDIA 合作頁面，在這裡您可以瞭解如何：

- **善用合適的專業知識。**對於基於 NVIDIA 的自訂 AI 和邊緣部署，我們擁有經驗豐富的諮詢和專業服務團隊，可以協助您充分利用 HPE 雲端採用架構。
- **最佳化 IT。**快速運轉容器化 AI 和 ML 環境，減輕資料科學生態系統的管理負擔，以便您可以輕鬆部署資源和容量，實現更好的業務成果。
- **更快實現價值。**精簡各業務單位的 AI 工作負載，以提高效能，協助資料科學家、開發人員、IT 團隊和研究人員將重心放回到解決方案建置和洞察收集上，從而加快實現價值的步伐。
- **僅按需進行支付。**輕鬆應對 AI 和 ML 工作負載的變化，並透過基於依使用付費模式的主動式容量管理來降低基礎架構成本。

* 可能設有最低容量或預留容量規定

如需瞭解詳情，請造訪以下網站：

HPE ProLiant 伺服器

HPE ProLiant AI 伺服器

線上交談 (銷售)

a50008742ZHT, 修訂版 1

Hewlett Packard Enterprise